SOUTH ASIAN LANGUAGE REIVEW
VOL. V, NO. 1, JANUARY 1995

TOWARDS THE COMPILATION OF A THESAURUS FOR MODERN TAMIL

S. RAJENDRAN

Introduction

A *thesaurus* in its widest contemporary sense is a classification of words by concepts, topics or subjects; it need not contain synonyms, and the fact that some items in a given class may be synonymous is coincidental. A *synonym dictionary*, by contrast, deals with word groups, and does not constitute a conceptual classification system of the kind exemplified by Roget's THESAURUS OF ENGLISH WORDS AND PHRASES (1852). This is a very crude distinction: thesauruses frequently contain synonyms, and synonym dictionaries are often thesauric, if the groups in them are large and cross-referenced and reflect a wide interpretation of synonymy.

Tamil lexicographical tradition begins with metrical works superficially of the thesaurus type (நிகண்டு *nikaṇṭu*) (see James, this volume). However, these compilations do not fully satisfy the criteria one would wish to establish for a *thesaurus*, which is not simply a word-finding device but must be more powerful, namely a manifestation of an efficient and precise classification of the lexical items in a language, reflecting a formal model of the ways lexical items express meaning. In the *nikaṇṭus*, entries are not classified on the basis of lexical relations, or hierarchies, but merely on the basis of inclusion within topic specification. For example, in TIVĀKARAM, the names of celestial/atmospheric entities such as *sun*, *moon*, *star*, *sky* and *cloud*, and abstracts such as *time*, *period*, *day*,

night, month, year are given along with the names of supernatural powers or beings, reflecting the socio-cultural belief patterns of the period of composition of this *nikaṇṭu*. In Sivasubrahmanya Kavirayar's NĀMATĪPAM (1810) along with the synonyms denoting 'elephant' and its parts, there occur *inter alia* the items denoting diseases, characteristics and ornaments associated with elephants.

The antiquity of the extant *nikaṇṭus* means that many lexical items are now obsolete as far as modern Tamil is concerned. For example, among the synonyms given for *donkey* and *camel* even in the comparatively recent NĀMATĪPAM (stanza 230), only கழுதை and ஒட்டகம் are in use now:

காளவாய் வேசரிகோ கத்திரிக ருத்தபமே
நீள்கரம்வா லேப நெடுங்கழுதை / தாழ்வலிதா
மத்திரிதா சோக மிரவணநெ டுங்கழுத்த
லித்துணையு மொட்டகமா மே

Included in this stanza can be seen such names for *donkey* as: காளவாய், வேசரி, கோகு, அத்திரி, கருத்தபம், கரம், வாலேபம் and கழுதை; and for *camel*: அத்திரி, தாசோகம், இரவணம், நெடுங்கழுத்தல் and ஒட்டகம் (Vaiyapuri Pillai 1930: 66). The *nikaṇṭus* evolved as works whose intention was to facilitate the development of word-power by the memorisation of lexical items as synonymous or homonymous denotative sets and group names, and to complement grammatical knowledge gained from the classical grammars. The *nikaṇṭus* thus reflect meaning as capable of independent analysis without further lexical reference, whereas it is the notion of the meaning of a lexical item being defined by its relations with other items, determining the structure of the vocabulary of a language, which underpins the construction of a thesaurus (Cf. de Saussure 1916; Ullmann 1959: 157).

Lyons (1977) discusses in detail the semantic structure of the vocabulary of a language, in particular from the points of view of syntagmatic and paradigmatic relations. The latter are those into which a linguistic unit enters through being contrasted with or substitutable by other similar units in a given environment, whereas the former are the relations that a unit contracts by virtue of its co-

occurrence with similar units. Thus, in வெள்ளை ரோஜா 'white rose' and சிவப்பு ரோஜா 'red rose', வெள்ளை 'white' and சிவப்பு 'red' are in a paradigmatic relation with each other, while each is in a syntagmatic relation with ரோஜா. It is the paradigmatic relations amongst lexical items that are important to us here, relations such as those of *synonymy, hyponymy, compatibility* and *incompatibility*. Lyons, who discusses (1977: 261) the complementarity that exists between paradigmatic and syntagmatic relations, is of the opinion that any satisfactory theory of lexical structure must incorporate both these.

Componential analysis

Componential analysis, a procedure by which the meanings of lexical items are analysed in terms of their distinctive features or components, serves to help in the understanding of lexical relations. Accordingly, *ideal synonyms* are those which share fully the semantic components of their meanings; *hyponyms* are those which share the same semantic components as their superordinate terms; *compatible items* are those which share certain semantic components and possess certain other contrasting components; lexical items which do not share any components are *incompatible*.

Synonymy

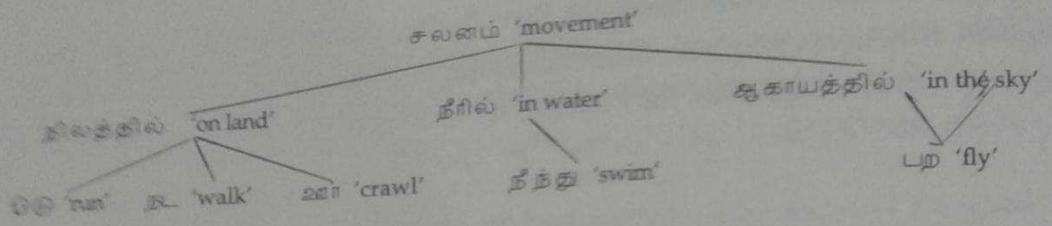
Ideal synonyms are hard to find: most so-called synonyms are quasi-synonyms, thus overlapping synonymy with compatibility. The diglossic nature of Tamil offers a greater possibility than otherwise for the occurrence of absolute synonyms, determined solely stylistically. Such synonymous items may differ in the emotive connotations they carry in addition to their cognitive meanings: வகை 'means, wherewithal', for example, is marked for contempt, whereas its "synonym" வழி is unmarked in this respect: cf. அவன் ஒரு வகைக் கில்லாதவன் and அவன் ஒரு வழியும் இல்லாதவன் 'he is without any

means', where the former connotes the speaker's contempt for the referent's situation. Synonyms may also differ according to collocational criteria: சிவப்பு 'red' has a wider range than, for example, its 'synonym' சிவலை whose range is more restricted (e.g. to காளை 'ox' or பசு 'cow').

Hyponymy

Hyponymy holds between a more specific or subordinate item (the *hyponym*) and a more general or superordinate item, and can be exemplified by such pairs as: பசு 'cow' : மிருகம் 'animal'; தாமரை 'lotus' : பூ 'flower'; and வில் 'sell' : கொடு 'give'. There does not always exist a superordinate for a class of co-hyponyms.

Hyponymy involves *entailment*, definable in terms of unilateral implication: இது ஒரு தாமரை 'this is a lotus' entails இது ஒரு பூ 'this is a flower', but not *vice versa*. Hyponymy also needs to be specified for immediacy, which is a transitive relation in a hierarchical configuration: thus பசு 'cow' is the immediate hyponym of பாலூட்டி 'mammal', which in turn is the immediate hyponym of மிருகம் 'animal': பசு 'cow' is thus also a hyponym of மிருகம் 'animal'. Hyponomy, for Lyons (1977: 294) is "a paradigmatic relation which rests upon the encapsulation in the hyponym of some syntagmatic modification of the sense of the superordinate lexeme". Generally, co-hyponyms of the same superordinate term will contrast by one or more semantic components: வில் 'sell' and வினியோகி 'distribute', which are co-hyponyms of கொடு 'give' can be contrasted by the additional semantic components they possess, apart from the common semantic components they share with கொடு 'give'. The relation of hyponymy reflects the hierarchical structure of the lexis and of the particular fields within the lexis; and the hierarchical ordering of lexical items can be represented formally thus:



The fact that the lexical items of a language may be classified into a number of parts of speech would appear to preclude the ordering of lexis in terms of hyponymy and superordinates. It is difficult to assign different parts of speech as hyponyms of a superordinate.

> If we include quasi-hyponymy with hyponymy as a relation in terms of which vocabularies are structured hierarchically, the hypothesis that the vocabulary in all languages is structured hierarchically under a relatively small set of lexemes of very general sense is rather plausible. It is a hypothesis, however, which is difficult to evaluate on the basis of the evidence that is at present available. (Lyons 1977: 299)

Conventional dictionaries depend on synonymy and antonymy to describe meanings of lexical items, whereas conceptual dictionaries or thesauruses rely on hyponymy and quasi-hyponymy.

> Indeed, it is hard to conceive of any language operating satisfactorily in any culture without its vocabulary being structured in terms of the complementary principles of hyponymy and contrast; and the descriptive work that has been done on various areas of the vocabulary in particular languages appears to support this conclusion. (ibid.: 301)

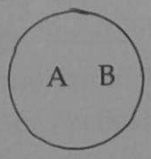
Lyons has taken up incompatibility (which includes antonymy as a special case) and hyponymy as fundamental relations to determine the structure of the vocabulary of a language and has kept aside synonymy as unimportant on the assumption that synonyms in language are redundant, since they can be removed without altering its basic structure. Jones (1986), on the other hand, argues that synonymy is more fundamental to other relations and attempts the semantic

classification of vocabulary of a language on synonymy. Nida (1975b) relies on *overlapping* (= synonymy), *contiguity* (the basic relation, a liberal extension of synonymy), *inclusion* (= hyponymy) and *complementation* (= lexical opposition). Cruse (1986: 86–87) establishes four *congruence relations*

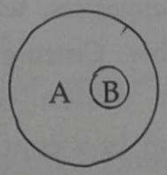
> for establishing a fundamental group of sense relations ... [and] also for defining a set of systematic variants applicable to virtually all other paradigmatic sense relations

in the following frameworks:

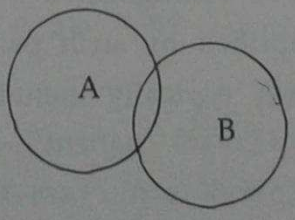
- *identity*: class A and class B have the same membership



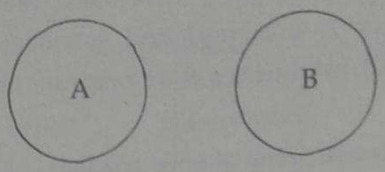
- *inclusion*: class B is wholly included in class A



- *overlap*: class A and class B have members in common but each has members not found in the other



- *disjunction*: class A and class B have no members in common



Fundamental to these relations is the question of whether the items involved overlap, exactly or inclusively or partially, or not. This wider sense of overlap may be associated with the synonymy, hyponymy (keeping aside partial overlapping, under contrast) and contiguity (separate from overlapping and inclusion) of Jones, Lyons and Nida respectively.

Lexical opposition

Antonymy, which is often considered as opposite to synonymy, relies on the lexical relation of incompatibility. Lyons reserves the term *antonymy* for a narrow sense, and describes several different kinds of lexical 'opposites'. For him, antonyms are always *gradable opposites*, which are distinguished from *complementaries*, which are not gradable. Cutting across this distinction of antonyms and complementaries is that drawn between *privative* and *equipollent* opposites: a privative opposition is a contrastive relation between two lexical items, one of which denotes some positive property and the other of which the absence of that property (e.g. உயர்திணை 'rational' : அஃறிணை 'irrational'); an equipollent opposition is a relation in which each of the contrasting lexical items denotes a positive property (e.g. ஆண் 'male' : பெண் 'female').

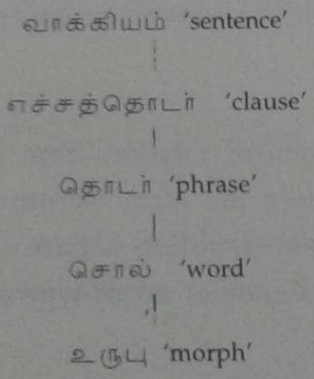
A further opposition to be distinguished from antonymy and complementarity is *converseness*, exemplified by such pairs as கணவன் 'husband' : மனைவி 'wife'. Converses typically include reciprocal social rôles (e.g. வைத்தியர் 'doctor' : நோயாளி 'patient'), kinship relations (e.g. அப்பா 'father' / அம்மா 'mother' : மகன் 'son' / மகள் 'daughter'), temporal relations

(e.g. முன்னே 'before' : பின்னே 'after'), spatial relations (e.g. மேலே 'above' : கீழே 'below') and complex relations (e.g. வாங்கு 'buy' : வில் 'sell').

In addition, there can be defined three further oppositions, those of *directional opposites* (e.g. சேர் 'arrive' : புறப்படு 'depart'; வா 'come' : போ 'go'), *orthogonal opposites* (e.g. perpendicularly opposite pairs such as வடக்கு 'north' : கிழக்கு 'east' and மேற்கு 'west'; and கிழக்கு 'east' : தெற்கு 'south' and வடக்கு 'north') and *antipodal opposites* (e.g. diametrical opposites such as வடக்கு 'north' : தெற்கு 'south'; கிழக்கு 'east' : மேற்கு 'west'). The vocabulary of kinship and colour can also manifest the principles of orthogonal and antipodal oppositions in many ways.

There are many multi-member lexical sets in a language whose lexical relations can be described as *incompatibility* denoting non-binary contrasts as opposed to binary contrasts (e.g. ஞாயிறு 'Sunday', திங்கள் 'Monday' ... சனி 'Saturday'). Various kinds of ordering are found in multi-member sets of incompatibles, and such sets may be serially or cyclically ordered. In the former there are two outermost members and all other lexical items in the set are ordered between others; in the latter, every lexical item is ordered between two others:

Serial

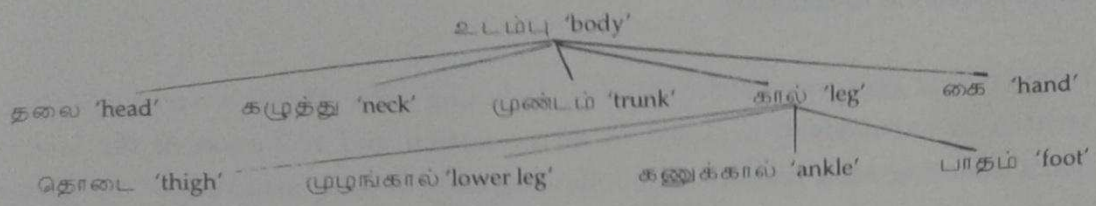


Cyclical

Among serially ordered sets, *scales* may be distinguished from *ranks* according to whether the constituent items are gradable or not. The ordering in scales in terms of incompatibility is typically less strict than in ranks (e.g. மிகநன்று 'excellent' : நன்று 'good' : திருப்தி 'satisfactory' : மோசம் 'bad' : படுமோசம் 'very bad'), whereas ranks exhibit the principle of strict serial ordering (e.g. ஒன்று 'one', இரண்டு 'two' ... நூறு 'hundred' ... ஆயிரம் 'thousand' ...). The best examples of cyclical sets or cycles can be found amongst lexis denoting units of time, e.g. வசந்தம் 'spring' : கோடை 'summer' : இலையுதிர் 'autumn' : குளிர் 'winter'; ஞாயிறு 'Sunday', திங்கள் 'Monday' ... சனி 'Saturday'.

Meronymy

A somewhat different hierarchical relation from hyponymy is the part-whole relation. Cruse (1986: 159) uses the term *meronymy* to denote "the semantic relation between a lexical item denoting a part and that denoting the corresponding whole". Meronymy is another major type of branching hierarchy:



There are numerous lexical items whose meaning cannot be specified independently of some part-whole relation of sense. For example, the meanings of செகண்ட் 'second', மினிட் 'minute' and மணிக்கூர் 'hour' cannot be stated clearly without reference to the meronymy within this lexical set. Further, although hyponymy can be distinguished from meronymy in principle, the distinction is blurred in the case of abstract nouns. The members of collective sets such as கால்நடை 'cattle', மந்தை 'herd', குடும்பம் 'family' can be viewed as having a meronymic relation with collective terms. Many such collectives can serve as superordinates in relation to sets of quasi-hyponyms, where there is an ambivalence with respect to the distinction of hyponymy and meronymy.

Meronymy forms the basis for several hierarchical classifications, and lexical items can be grouped under various heads based on this relation. Meronymy as a part-whole relation seems to exist between an individual lexical item and the lexical field which contains it, which is similar to the relationship existing between the lexical field and the total vocabulary of the language.

Semantic classification

The classifications given in NĀMATĪPAM (1810) and in Nida (1975a) may serve as examples for the elcidation of problems involved in the semantics of contemporary Tamil lexis.

Classification in NĀMATĪPAM

NĀMATĪPAM comprises 808 quatrains, including some 12,000 lexical items. It is written in *sutra* style, to assist in memorisation and recitation of the lexical items denoting each referent. The following embodies the structure of the lexical classes of this *nikaṇṭu*:

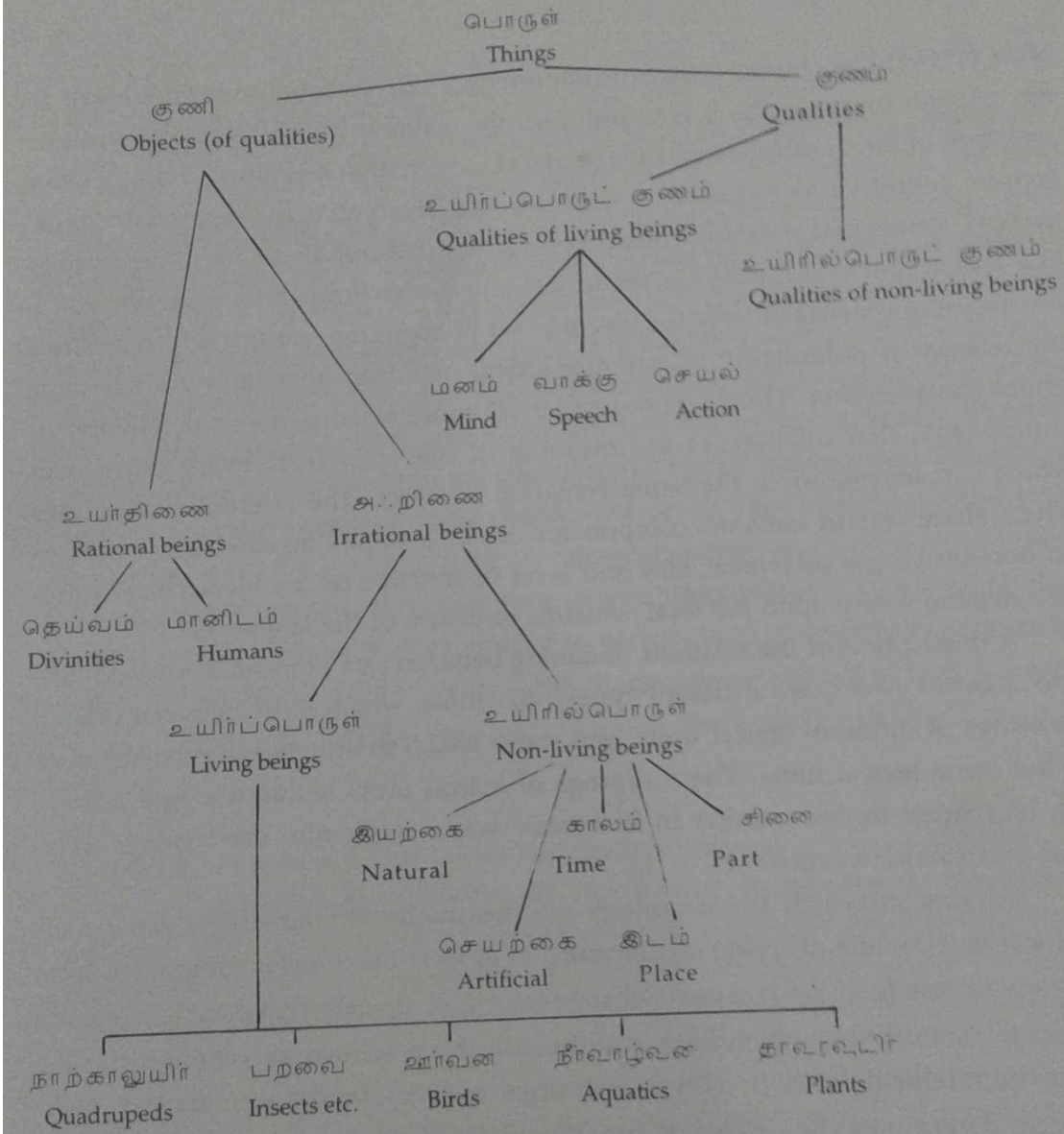
I Rational creatures:
1 divinities
2 human beings

II Irrational creatures:
1 quadrupeds
2 birds
3 insects etc. ("class of crawling beings")
4 fish and reptiles
5 plants

III Irrational non-living beings:
1 natural things
2 artificial things
3 place
4 time
5 part

IV Qualities:
1 living creatures: mental qualities
2 living creatures: communicative qualities
3 living creatures: qualities of action
4 qualities of non-living beings

The macrostructural division of the vocabulary into four major chapters, and their sub-divisions, characterise the *nikaṇṭu* as a thesaurus. The following is the hierarchical classification of NĀMATĪPAM, as given by Vaiyapuri Pillai (1930: v):



This classification, although a refinement of similar schemata in earlier *nikantus*, still does not accurately reflect the lexical structure of Tamil.

Classification by Nida

Nida (1975a) has chosen componential analysis of referential meaning, based on the relation between the lexical unit and the referent, to build up the semantic structure of the vocabulary of a language. His referential meaning is based on two factors: cognitive as opposed to emotive, and referential as opposed to grammatical. Referential meaning consists of a structured bundle of cognitive features of a lexical unit which facilitate the lexical unit designating all its referents, that is the meaning consists of only the essential set of cognitive features which separate the referential potentiality of any one lexical unit from that of any other units under consideration. Thus, one must identify those essential semantic features of components that distinguish the meaning of one item from every other item which can be placed in the same semantic territory. The referential meanings which share certain semantic components can be grouped together into a semantic domain whose relevance, size and level of function in the hierarchical structure depend solely upon the total semantic structure of the language.

A description of the relations obtaining between sets of meanings has to take into account two quite different types, viz. those which exist between related meanings of different lexical units, and those which exist between related senses of the same lexical units. The meanings of lexical units within a single domain can be related to one another in four basic ways: *contiguity, overlapping, inclusion* and *complementation.*

The sets in which the meanings are maximally similar at the same hierarchical level, while differing significantly in at least one crucial component form *contiguous* sets (e.g. நாற்காலி 'chair' : பெஞ்சு 'bench' : முக்காலி 'stool' : சோபா 'sofa'). Sets which have traditionally been known as synonyms at the same hierarchical level in which meanings overlap (and the extent of overlapping determines the degree of synonymy) form *overlapping* sets (e.g. கொடு 'give' : அளி 'bestow'). *Included* sets are those in which meanings are related to one another in hierarchical series and the included meanings have at least one more diagnostic component than the meanings in which they are included (e.g. மிருகம் 'animal' : பாலூட்டி 'mammal' : பூனைக் குடும்பம் 'family

felidae : புலி 'tiger'). *Complementary* sets of meanings are related in three ways: *opposites* (e.g. பெரிது 'big' : சிறிது 'small'), *reverses* (e.g. கட்டு 'tie' : அவிழ் 'untie') and *converses* (e.g. வாங்கு 'buy' : வில் 'sell').

The different meanings of the same lexical unit can be related by four principal relational types, viz. *derivation, replacement, figurative extension* and *peripheral clustering*. *Derivation* involves the incorporation of the semantic components of the base into the derived meaning belonging to a different semantic domain (e.g. அழகு 'beauty' : அழகான 'beautiful'); *replacement* involves the substitution of at least one semantic component in the base by another different but related component without a major shift in the semantic domain (e.g, ராஜா 'king' : ராணி 'queen'); *figurative extension* involves totally different semantic domains with relations between the base and the figurative meanings established by means of some supplementary component or reinterpreted diagnostic component (e.g. நரி 'fox' : தந்திரசாலி 'deceptive and/or crafty person'); *peripheral clustering* involves the clustering of meanings which are related to their central meaning (if there is one) and/or to each other by the presence of certain common components and by links in the diagnostic com-ponents (e.g. கெட்டி 'firmness, thickness' in கெட்டியான தோல் 'thick skin' and கெட்டியான பாகு 'thick syrup').

> An analysis ... leads one to recognize four basic semantic classes of symbols. ... (1) "objects," to be found in such words as *house, tree, leaf, sun, hill, man, rock, wall* (but not necessarily in all occurrences of these words), (2) "events," in such words as *run, walk, cut, talk, jump, go, sit, squirm*, (3) "abstracts," in such words as (a) *red, blue, small, many, one, two* (abstracts of objects), (b) *swift, slow, once, twice, often* (abstracts of events), (c) *hard* (e.g. *hard rock, worked hard*), *soft* (e.g. *soft coat, soft job*) (abstracts of either objects or events) or (d) *very* (e.g. *very many*), *too* (e.g. *too small*), *so* (e.g. *so swift*) (abstracts of abstracts), and (4) "relationals," which serve to relate various objects, events and abstracts, e.g (a) object to object: *John and Bill, the men in the house, the book on the table*; (b) object to event: *the men in flight, the barking of the hounds*; (c) event to event: *run and jump, won in swimming*; (d) objects to abstract: *an object of beauty, the color of a rose*; (e) abstract to event: *adept at winning, good for eating*; (f) abstract to abstract: *bad and lazy, beauty of*

> *holiness*. These four basic semantic classes not only exist in all languages but appear to be fundamental to the analysis of meaningful relationships between all types of lexical forms.
> (Nida 1975b: 152)

The following is an abbreviation of Nida's (1975a: 178–186) classification, giving only a selection of the major subdomains adopted:

I Entities
- A Inanimate
  - 1 natural
    - a geographical
    - b natural substances
    - c flora and plant products
  - 2 manufactured or constructed entities
    - a artefacts (non-constructions)
    - b processed substances; foods, medicines and perfumes
    - c constructions
- B Animate entities
  - 1 animals, birds, insects
  - 2 human beings
  - 3 supernatural powers or beings

II Events
- A Physical
- B Physiological
- C Sensory
- D Emotive
- E Intellectual
- F Communication
- G Association
- H Control
- I Movement

J Impact
K Transfer
L Complex activities, involving a series of movements or actions

III Abstracts
A Time
B Distance
C Volume
D Velocity
E Temperature
F Colour
G Number
H Status
I Religious character
J Attractiveness
K Age
L Truth–falsehood
M Good–bad
N Capacity
O State of health
etc.

IV Relationals
A Spatial
B Temporal
C Deictic
D Logical
etc.

This classification is based on referential meanings, and it is not possible to obtain one-to-one correspondences between the semantic domain of classes and grammatical classes; there are, of course, certain parallels, however,

> since on some level of the deep structure entities tend to be represented by nouns, events by verbs, abstracts by qualifiers, and relations by a number of different features: particles, affixes of case, word order, etc. (Nida 1975a: 176)

There is evidently a clear logic governing such a classification, though it is not difficult to construct an alternative schema, whilst retaining a similar line of thought with reference to the vocabulary of a language. Nida's universal semantic classification can be adapted for Tamil without drastic alteration, though there are some problems in wholesale adoption.

Adapting Nida's classification for Tamil

It is sometimes difficult to compartmentalise items of Tamil vocabulary to fit into the semantic framework prepared by Nida. There can often be a vagueness in deciding whether a particular lexical item does or does not belong to a particular semantic domain into which Nida has classified referential meanings. The division of entities into ANIMATE and INANIMATE, for example, needs reformulation. Tamil grammatical and lexicographic traditions divide nouns into classes of RATIONAL and IRRATIONAL BEINGS, the former including divinities and human beings and the latter all other kinds of creature. (Moreover, Nida includes plants as INANIMATES, whereas Tamil tradition classifies flora as IRRATIONALS (NON-HUMAN, INANIMATE). This division has a direct bearing on the morpho-syntax of the language, in particular the collocational occurrence of third person pronominal forms, gender distinction and the personal terminations of finite verbs. Thus, அவன் 'he', அவள் 'she', அவர் 'he/she (respectful)' are possible only with reference to rational beings, அது 'it' and அவை 'they' to irrational beings. Accordingly, we have the finite forms of the verbs in examples such as: வந்தான் "came-he", வந்தாள் "came-she", வந்தார் "came-he/she (respectful)", வந்தது "came-it", வந்தன "came-they (irrational)".

Also, traditionally plural nouns of irrational beings are optionally marked by the suffix -கள் (e.g. பசு / பசுக்கள் 'cow / cows'), whereas the plurals of

rational beings are marked by -அர் and its alternants, e.g. தூதர் 'messenger', தோழியர் 'a lady's maids', துறவோர் 'ascetics'. In modern written Tamil, although -கள் is used to mark the plurals of both rational and irrational beings, e.g. பசுக்கள் 'cows', பெண்கள் 'women', it has almost disappeared from use as a plural marker for irrational beings in the spoken language (e.g. சில பசு 'a few cows'). Even the contrast between the personal terminations -அது 'third person irrational singular' and -அன 'third person irrational plural' is being lost in spoken Tamil (e.g. சில பசு வந்தது 'a few cows came').

Adaptation for Tamil of Nida's classification of EVENTS gives rise to several problems. For example, he has grouped *break, crush–smash, cut–stab, beat, kill, ruin–destroy* etc. under the subclass IMPACT, without, however, elucidating whether the items denoting the process or change of state such as *being broken, being crushed* or *smashed, being cut, being ruined* or *destroyed* etc. are also to be included here or (also) under PHYSICAL EVENTS. Similarly, the placing of events of FOOD PROCESSING under COMPLEX ACTIVITIES INVOLVING A SERIES OF MOVEMENTS OR ACTIONS is inconsistent with regard to the placing of the event of process of *being cooked*. It appears from his principle of relating derivational meaning to base that the process of change of state (*being broken* etc.) has to be placed alongside events of process–action (*break* etc.). However, the problem here is which is to be considered as the base: the process of change of state or the process–action. Nida's other classes such as ASSOCIATION, CONTROL, MOVEMENT and TRANSFER demonstrate that he has taken the volitional process–action as basic to process.

Elsewhere (Rajendran 1978, 1981) I have preferred to consider CHANGE OF POSITION (i.e. movement) and CHANGE OF STATE as the foremost events common to animate and inanimate entities, with CHANGE OF POSSESSION (i.e. transfer), SENSORY, EMOTIVE, INTELLECTUAL and COMMUNICATIVE EVENTS and CALLS (e.g. animal calls) as belonging to animate entities, along with a miscellaneous group including ASSOCIATION, COOKING, CONTROL and a number of others. I have taken events of process as more basic than events of process–action in Tamil because many of its transitive verbs can be seen to be derived from their respective

intransitives by the process of transitivisation, e.g. நசுங்கு 'be crushed' > நசுக்கு 'crush'. A problem arises in that many events are able to be stated as belonging to a class of CHANGE OF STATE and having the corresponding events of IMPACT, ASSOCIATION, CONTROL and COMPLEX ACTIVITIES in the form of transitive verbal counterparts of intransitives. A subclass of events of CHANGE OF STATE, with cross-reference to events of ACTION–PROCESS under IMPACT, ASSOCIATION, CONTROL and COMPLEX ACTIVITIES would have to be included in the semantic specification, but to avoid repetition intransitive forms are better given under the respective transitive forms. This is the procedure I have adopted in the thesaurus work I have already completed (Rajendran 1991).

The class of abstracts also poses certain similar problems. Nida has a subclass TEMPERATURE, under which he has listed abstracts such as *hot, cold, lukewarm* etc. In Tamil, for example the verbs சூடு 'be hot' and குளிர் 'be cold', denoting sensation by contact, have corresponding nominal forms சூடு 'heat' and குளிர் 'cold'. சூடு and குளிர் can be classified as SENSORY (under EVENTS), with a cross-reference to TEMPERATURE (under ABSTRACTS). Many such relations between entities and abstracts, and events and abstracts, have thus been tackled, adopting Nida's principle of relating inflectional and derivational meanings to their respective bases. For example, the relations between நண்பன் 'friend' and நட்பு 'friendship', பெற்றோர் 'parents' and பெறு 'give birth', உரை 'tell' and உரை 'commentary' can be shown by cross-reference.

Inflectional and derivational bases

Nida does not include separate classes for performers of functions; for example, he treats *ruler, builder* etc. as derivatives of the underlying EVENTS *rule, build* etc. For him, *beautiful* and *beautifully* are to be entered as ABSTRACTS, as their referential meaning is similar; and *friendly* is entered alongside *friend*, even though the former denotes an abstract quality. He prefers to classify *carpenter* under COMPLEX ACTIVITIES along with *builder, worker, labourer* etc. even

though *carpenter*, unlike the other examples, does not have a base denoting the corresponding event. Following this schema uncritically for Tamil, one would be forced to classify together as EVENTS many verbal derivatives which are formed from verbs by the addition of derivative suffixes changing the basic verbal meanings into derived meanings denoting agents, instruments etc. (e.g. காதலன் 'lover' < EVENT காதலி 'love'; துடைப்பம் 'broom' < EVENT துட 'rub'). By doing so, although one can demonstrate the transfer of semantic components of the respective events to agents, instruments etc., the membership assigned to various domains will be significantly disturbed. For example, non-verbal derivative agents such as தச்சன் 'carpenter' or பொற்கொல்லன் 'goldsmith' would have to be placed alongside verbal derivatives such as ஓட்டுநர் 'driver' (< ஓட்டு 'drive') or நடத்துநர் 'conductor' (< நடத்து 'conduct'), whereas because Nida has a subdomain TOOLS under ENTITIES, non-verbal derivatives such as வாள் 'sword' or கோடாலி 'axe' would have to be entered separately from verbal derivatives such as கத்தரி 'scissors' (< கத்தரி 'cut') or சுரண்டி 'scraper' (< சுரண்டு 'scrape'). A similar dilemma obtains with respect to items denoting taste, such as இனிப்பு 'sweet', கசப்பு 'bitter' and புளிப்பு 'sour', which have to be listed along with their respective EVENTS under SENSORY 'taste sweet', கச 'taste bitter' and புளி 'taste sour' respectively) or separately along with ABSTRACTS under the class TASTE (by analogy with COLOUR).

Nida's reason for placing derived forms alongside their bases is to assign componential features to the derivatives from the base, thus avoiding duplication. In general, however, to include derived forms under the class to which they belong semantically and furnish them with a cross-reference for the bases from which they are derived seems to be an intuitively more satisfactory solution with respect to Tamil.

Superordinate terms

The availability of a superordinate term which can cover a semantic domain facilitates the assigning of membership to the domain. For example, விலங்கு 'animal', in a generic sense comprises all quadrupeds, but as a technical term it can cover all living beings other than plants. Often there may be no generic form to include a group of items, or a form may be both generic and non-generic. The non-availability of a generic term can be managed by a number of techniques: a specific item can be generalised as a superordinate term, e.g. the specific பூனை 'cat' is generalised to cover cats, lions, tigers etc., as பூனைக் குடும்பம் 'family *felidæ*'; a superordinate term can be a descriptive phrase, as in நீரில் வாழும் உயிர்கள் 'aquatic beings' or பூகோளம் சம்பந்தப் பட்டவை 'things related to geography' for 'geographical'. On the other hand, all plants can be captured by the term தாவரங்கள் 'plants' and all colour terms by நிறங்கள் 'colours'. Often, a term used as a superordinate to denote a domain or a class has to be accompanied by a definition, e.g. பருப் பொருட்கள் 'entities' may need an explication such as 'objects having concrete shape or which can occupy space'.

Structuring lexical items into domains

The hierarchical structuring of lexical items and the establishment of boundaries between domains can sometimes pose difficulties. One cannot always expect a sharply defined and dichotomously classified distinction. It may be relatively simple to decide upon those items which form the core of a domain, but there may be many borderline cases where the decision of boundary establishment becomes problematical. For example, ஓடு 'run', நட 'walk', நகர் 'move' etc. can be easily assigned as EVENTS denoting MOVEMENT, but the meanings of verbs such as ஒழுகு 'leak, flow', கசி 'ooze', வடி 'percolate, flow', வழி '(over)flow' and ததும்பு 'overflow' can belong to the class PHYSICAL EVENTS INVOLVING LIQUIDS as well as MOVEMENT. Problems of this sort can be solved

> only in terms of the total picture of semantic relations. The justification for assigning a meaning to a particular domain, or for relating domains into larger groupings, depends primarily upon two factors: (1) the extent to which certain components are shared and (2) the relative importance of these components. But the relative importance of components can be known only in terms of the extent to which certain classifications provide a basis for "rules" with greater explanatory power, in the sense that they can "predict" precisely the semantic features of such language functions as naming, paraphrase, and collocation.
>
> (Nida 1975a: 193)

Polysemy gives rise to further problems: நட, for example, has two principal meanings: 'walk', denoting the actual movement by means of limbs, and 'happen', the abstract movement of function, performance, work, business, life, age etc. (cf. Rajendran 1978: 95). Thus நட needs to be split into நட$_1$ 'walk' and நட$_2$ 'happen', the actual MOVEMENT of 'walk' and the ABSTRACT MOVEMENT of an EVENT.

Scientific versus *folk taxonomies*

There is often a clash between scientific taxonomy and folk taxonomy, particularly in such domains as those of flora and fauna. As Tamil is aiming at equipping itself for scientific use, preference may be felt for a scientific classification in a thesaurus of modern Tamil, over that enshrined in folk taxonomy, which reflects the culture. But compromises are inevitable. For example, தென்னை 'coconut', பனை 'palm' and வாழை 'plantain' are traditionally மரங்கள் 'trees', whereas scientifically they are புற்கள் 'grasses'.

Classifications in the *nikaṇṭus* offer us many instances of such folk taxonomies, which have to be modified from a scientific point of view. For example, NĀMATĪPAM has included the names of snakes, ants, scorpions, spiders, centipedes, worms, lizards, woodworms, termites, bellyworms, crickets etc. along with their generic name ஊர்வன 'crawling beings' under ஊருயிர் பொருள் வர்க்கம் 'class of crawling beings'. Modern scientific classificatory schemata

would adopt ஊர்வன for 'reptiles', கணுக்காலிகள் for 'arachnids', பூச்சிகள் for 'insects' and புழுக்கள் for 'worms'. Similarly, the names of fish (which include திமிங்கிலம் 'whale'), frogs, crabs, leeches, crocodiles, tortoises and molluscs given under நீரில் வாழும் பொருள் வர்க்கம் 'class of aquatic beings' in the *nikaṇṭu* have to be reclassified adopting a scientifically oriented taxonomy.

One can often find analogies between the structure of a culture and the semantic structure of a language. The classification of celestial and atmospheric phenomena along with supernatural beings and astrologically believed entities along with astronomical phenomena in the *nikaṇṭus*, for example, reflects the cultural thinking of the periods in which they were compiled. The relations between features of semantic structure and culturally significant attitudes and beliefs are transient and subject to change, as views concerning reality alter over time. Pre-scientific orientation can lead to scientific orientation. The controversy between the conscious realisation and the unconscious tradition can be discerned, for example, in the use of two alternants for the denotation of the young of animate beings as குட்டி or குஞ்சு. In ordinary language, குட்டி can be used to denote the young of all animals, and குஞ்சு is used mainly for the young of birds. (There are a number of exceptional and overlapping cases due to dialectal and idiosyncratic differences, however.) But the scientific distinction between young born by delivery (குட்டி) and by hatching (குஞ்சு) cuts across the generalised conceptualisation of the ordinary usage of குட்டி, and demands குஞ்சு instead of குட்டி for the young of the *crocodile*, and குட்டி instead of குஞ்சு for the young of the *bat*. This represents a binary classification of animals into mammals and others.

What is expected is a scientifically principled modification of taxonomy rather than necessarily a technical scientific exposition. Modern scientific taxonomy excludes இரால் 'shrimp' from its traditional place in the class of fish, and classifies வௌவால் 'bat' not as a bird but as an animal. We would wish to place பூமி 'earth' and சந்திரன் 'moon' separately from நட்சத்திரங்கள் 'stars', and சூரியன் 'sun' along with 'stars', and to eliminate such items as

இராகு 'a mythical serpent believed to consume the sun and the moon and cause eclipses' and கேது 'a mythical celestial dragon believed to cause eclipses' from the class of astronomical phenomena. Preference for astronomy over astrology reflects a contemporary scientific approach to thesaurus construction.

In equipping itself to serve as a communicative medium for science, Tamil coins lexical items by loan translation from English as well as trawls its own internal sources. In the absence of a superordinate term for a group of items, for example, a hyponymous item may be used with an attribute. A specific item may be raised to the level of a superordinate for the purpose of scientific classification. For example, specific items such as பூனை 'cat', நாய் 'dog' or கரடி 'bear' are used as superordinate scientific classificatory epithets denoting the various families respectively பூனைக் குடும்பம் 'family *felidæ*', நாய்க்குடும்பம் 'family *canidæ*', கரடிக்குடும்பம் 'family *ursidæ*'. It is noteworthy that the *nikaṇṭus* used a similar procedure of raising a specific item to a superordinate: for example, விலங்கு, which denotes 'quadrupeds' in ordinary language, was used in the *nikantus* to denote a class of 'living beings other than plants and humans', and similarly, மரம் 'tree' was used as a superordinate term in மரப்பெயர்கள் 'names of plants'.

Part-whole relation

Meronymy was noted even by the early Tamil grammarians and lexicographers. *Nikaṇṭus* such as TIVĀKARAM gave names of parts along with their respective wholes, but NĀMATĪPAM has a separate section for meronymy or சினை முதல் 'part-whole', in which the names of a body and its parts are listed alongside related entities, e.g. stanza 565

தேகமுடல் காத்திரமெய் பூட்சிபுர மாதார
மாக முயிர்நிலைகா யஞ்சடலந் / தோய்குரம்பை
புட்கலங்க னம்புணர்ப்பு டம்புகடம் யாக்கைசடம்
விக்கிரகம் பூதியம்விம் பம்

Vaiyapuri Pillai (1930: 169) identifies the items in this stanza (சரீரத்தின் பெயர் 'names of the body') as: தேகம் 'body', உடல் 'body', காத்திரம் 'body', மெய் 'body', பூட்சி 'body', புரம் 'body', ஆதாரம் 'body as the receptacle for the soul', ஆகம் 'body as connected with transmigration or as assumed by a deity', உயிர்நிலை 'body as the seat of life, or station of the soul', காயம் 'body', சடலம் 'lifeless body', குரம்பை 'body as a frame', புட்கலம் 'fulness, body', கனம் 'heaviness, body', பணர்ப்ப 'body as composed of various parts', உடம்ப 'body', கடம் 'body', யாக்கை 'body as consisting of fibres and ligaments', சடம் 'body', விக்கிரம் 'body as an image', பூதியம் 'body' and விம்பம் 'form, image'. Cf. stanza 578:

அங்கிரிப தஞ்சலனந் தாள்சரணம் பாதமடி
பைங்கழலுங் காலாம் பரடுகர / டங்கரண்டை
காண்கணைங்கால் சங்கமுழங் கான்முழந்து சானுமுட்டா
மாண்குறங்கூ ருத்துடைக வான்

Vaiyapuri Pillai (1930: 170) identifies the items in this stanza (காலின் பெயர் 'names of the leg') as: அங்கிரி 'foot', பதம் 'foot', சலனம் 'limb (leg)', தாள் 'foot', சரணம் 'foot, especially of a deity or great person', பாதம் 'foot', அடி 'foot', கழல் 'foot', கால் 'leg, foot'; (கணைங்காலின் பெயர் 'names of the shank':) கணைங்கால் 'shank', சங்கம் 'shank'; (கரண்டைக்காலின் பெயர் 'names of the ankle':) பரடு 'ankle', கரடு 'ankle', கரண்டை 'ankle'; (முழங்காலின் பெயர் 'names of the knee':) முழங்கால் 'knee', முழந்து 'knee', சானு 'knee', முட்டு 'knee'; (துடையின் பெயர் 'names of the thigh':) குறங்கு 'thigh', ஊரு 'thigh', துடை 'thigh', கவான் 'thigh'.

It is uncontroversial to list generic parts under a lemmatised generic whole, but the inclusion of specific parts under specific bases raises problems. For example, parts such as வேர் 'root', தடி 'trunk', கிளை 'branch', இலை 'leaf', பூ 'flower', காய் 'unripe fruit' or பழம் 'ripe fruit' listed under மரம் 'tree' pose little theoretical problem. However, this is not the case with specific parts such as மாம்பூ 'mango flower', மாங்காய் 'unripe mango', மாம்பழம் 'ripe mango', மாந்தளிர் 'tender leaf of mango' or மாஞ்செடி 'mango plant'. The availability

of sandhi-determined stem alternants (e.g. மாம்-, மாங்-, மாந்-, மாஞ்-) could be used in justification for the individual listings, but this will lead to over-repetition, and in any case there are numerous cases where sandhi rules are inoperative and the base form does not change, e.g. கொய்யா 'guava tree': கொய்யாப் பழம் 'ripe guava', கொய்யாக்காய் 'unripe guava' and கொய்யாச்செடி 'guava plant'. Stem alternants could be listed along with the lemmatised base, and cross-referred to the class of generic parts.

The situation is, however, further complicated by the process of *metonymy*, in which a part can stand for a whole or a whole for a part, e.g. முல்லை 'jasmine' or ரோஜா 'rose' can denote the respective plants as well as their flowers and in the same way, ஆரஞ்சு 'orange' or தக்காளி 'tomato' can denote the respective plants as well as their fruits. This is not always the case, though, as for example மா 'mango' and பலா 'jack' denote only the trees, with the phrasal forms மாம்பழம் and பலாப்பழம் denoting their respective fruits. Furthermore, there are certain forms which are exclusive to certain parts of wholes, e.g. மடல் 'leaf', ஓலை 'ola' and நுங்கு 'palmyra fruit' are specific with to reference to the palm, and தந்தம் 'ivory' and தும்பிக்கை 'trunk' have specific reference to the elephant. There are other cases where there are collocational restrictions regarding the formation of compounds, e.g. கன்று 'young plant' can occur with தென்னை 'coconut', வாழை 'plantain', மா 'mango' etc. to form compounds such as தென்னங்கன்று 'coconut seedling', வாழைக்கன்று 'plantain seedling' or மாங்கன்று 'mango seedling' etc., but with செடி 'small plant', compounds such as *தென்னஞ்செடி and *வாழைச்செடி do not occur although மாஞ்செடி 'mango plant' does. Rules and explanations for complexities such as these would be tedious, but listing derivatives alongside the specific bases may go some way towards providing a solution. However, innumerable compounds may have to be listed in some cases, e.g. under மா 'mango', one might have to list the following, in addition to those already discussed above: மாம்பலகை 'board of mango wood', மாந்தடி 'mango trunk', மாவேர் 'mango root', மாவிலை 'mango leaf' and so on. In this situation, one

solution would be to list a selection of names of parts accommodating the stem alternants, along with cross-references to the generic parts.

Selection

Stylistic alternants

The nature of Tamil as a diglossic language, with literary (H, or LT—see Schiffman, this volume) Tamil existing alongside the spoken (L, or ST) varieties, is well attested in the literature (cf. de Silva 1976; Schiffman 1978; Deiva Sundaram 1981; Britto 1986). So much prestige is attached to the H variety that it has over time become considered to be the 'pure' and the 'real' language, and the L varieties somehow inferior. H is not the same as 'pure' (although some scholars have suggested this) and H/L should, as Britto (1986: 324) notes, be defined, in terms of "acquisitional and functional superposition rather than in terms of prescriptive ideals such as purity and correctness". At the same time, a modern standard spoken variety is emerging as a kind of lingua franca amongst the Tamils, and in spite of discrepancies in this variety, there is an understanding amongst Tamil (at least amongst those who feel that there should be a standard spoken variety) that there are now two standard varieties of Tamil which have to be taken into account in any linguistic or lexicographic exposition, viz. H and SL (= "standard L"). 'Tamil', then, would connote at least these two varieties. As there are often distinct forms for the same lexical item, a dictionary or thesaurus which aims to represent contemporary Tamil should take these differences into account, perhaps either by giving the SL forms with H forms as sub-entries, or by lemmatising the H forms and offering phonological or other rules by which the conversion from H to SL can be made. Deiva Sundaram (1981: 44–53) has given a selection of such conversion rules to "change the forms in written variety [= H] into the forms of the spoken variety [= L]" (see also Shanmugam Pillai 1972a, 1972b; Sethu Pillai 1974; Asher 1985: 259–269; Britto 1986: 196–203).

Inflectional formations

A stem can be incremented by gender-number inflection as in e.g. மாணவன் 'male student' ~ மாணவி 'female student' ~ மாணவர் 'student'. It is not always the case that all three inflections are available for a stem: sometimes there may be just one (e.g. மனைவி 'wife'), sometimes two (e.g. கவிஞன் 'male poet', கவிஞர் 'poet (respectful)'). Each of these inflectional morphs has more than one alternant, the distribution of which is morphophonemically conditioned. The thesaurus would have to accommodate these sets of alternants within a lemma-used inflection, e.g.

-அன் ~ -ஆன் ~ -ஐயன் ~ -மகன் ~ -னன் ~ -ஆடி 'male singular'
(e.g. நீதிபதி 'friend', கலைஞன் 'artist', குடும்பஸ்தன் 'family man', புத்திசாலி 'clever man', கிராமத்தான் 'villager').

The plural marker -கள் creates a further dilemma of the same sort. -கள் is regularly added to the noun to denote plurality, and so normally there is no need to give the forms inflected for -கள் in the thesaurus. But, in certain cases a form inflected for the plural is preferred to the simple (singular) (e.g. பொதுமக்கள் 'common people' occurs more commonly than பொதுமக்களுள் ஒருவன் 'common people'). In certain cases, the plural inflection serves to offer a semantic contrast, e.g. அர்ச்சகர் 'priest, teacher' *vs* அர்ச்சகர்கள் 'priests, teachers' but also 'Brahmin temple priest'—because of this specific derived meaning, the inflected form has to be given in the thesaurus. In பிள்ளைகள் 'children, people', this form has to be taken as a lemmatised entry, because the -கள் inflection is attached to the bound stem பிள்ளை-, which has no independent existence: cf. பிள்ளை 'son'.

Inclusion of borrowings from English

Modern Tamil has adopted a large number of borrowings from English, e.g.

டைம் 'time', மினிட் 'minute', செகண்ட் 'second', பேப்பர் 'paper', பஸ் 'bus', ஐஸ் 'ice', செயர் 'chair', வாச்சு 'watch' (cf. Tucker 1986). Adherents of the purist movement in Tamil resist the inclusion of borrowings from English in any writing, including dictionaries and thesauruses. Purism, in the words of Annamalai (1979: 36), is "the opening of the native sources and closure of the non-native sources for the enrichment of a language", and in the modern world, it is argued, non-native sources should not be closed off for lexical creation. Tamil is in urgent need of lexical items to express contemporary concepts, particularly in the realms of scientific and technical knowledge, and the rich source of lexis in these fields in English is consistently tapped by Tamil, though without the approval of purists. In the absence of any empirical studies concerning the qualitative and quantitative statistics of borrowing from English into Tamil, items have to be included in a modern thesaurus on an impressionistic basis only. Tamil dictionaries are, moreover, of little help in this regard, as their compilers have in general eschewed the inclusion of borrowings from English.

There are, however, principles which transcend an impressionistic view. The regular use of a particular borrowed item by otherwise monolingual Tamil speakers must give that item a different status from those used deliberately by bilingual speakers when they mix codes. For example, borrowed words such as பஸ் 'bus' and பென்சில் 'pencil' are used extensively even by monolingual speakers of Tamil, and thus have to be included in a thesaurus, whereas examples such as டேபிள் 'table', புக் 'book', ∴பான் 'fan' and புரபசர் 'professor' do not pass muster according to this criterion, as the readily available (not necessarily native-) Tamil equivalents (e.g. மேஜை 'table', புஸ்தகம் 'book', விசிறி 'fan', பேராசிரியர் 'professor') are more extensively used by monolingual native speakers.

Inclusion of phrases and compounds

The compiler of a thesaurus has to decide upon the status of phrases and compounds for inclusion in the work. Obvious instances of items to be included

would be e.g. compounds with நல்ல 'good', where an idiomatic or exocentric meaning occurs: cf. நல்ல பாம்பு 'good snake' and நல்லபாம்பு 'cobra', or நல்ல நெய் 'good ghee' and நல்லெண்ணெய் 'gingely, or sesame seed, oil'. On the other hand, compounds formed by a highly productive mechanism would not need to be included, e.g. கோழி முட்டை 'hen's egg', வாத்து முட்டை 'duck egg', பாம்பு முட்டை 'snake egg' etc., whereas compounds whose formation is generally not productive do have to find a place in the thesaurus, e.g. compounds with உரை 'utterance' such as: விரிவுரை 'lecture', பொழிப்புரை 'summary of meaning, synopsis', பதவுரை 'verbatim meaning' and தெளிவுரை 'clear description'.

Inclusion of proper names

NĀMATĪPAM, in common with a number of early *nikaṇṭus*, gives various names for each Hindu god and goddess: Siva, for example, is given 161 names in ten stanzas (9–18), Vishnu, 120 in eight stanzas (44–51) and Parvati, 99 in six stanzas (19–24). Sadasivam, a modern lexicographer, has prepared the following hitherto unpublished compilations: A DICTIONARY OF EPITHETS AND NAMES OF LORD MURUGA (with 12,000 entries), A DICTIONARY OF EPITHETS AND NAMES OF LORD VISHNU (with 20,000 entries), A DICTIONARY OF EPITHETS AND NAMES OF GODDESS UMA (with 6,000 entries), A DICTIONARY OF EPITHETS AND NAMES OF LORD SIVA (with 30,000 entries) and A DICTIONARY OF EPITHETS AND NAMES OF GOD (with 5,000 entries) (Vadivelan 1980; James 1991: 227–228). It is evident that specialised lexical lists of such proportions cannot find a place in a thesaurus if it is to remain readily accessible; furthermore, the inclusion of proper names of literary and famous figures, writers, mythological characters etc. etc. would change the character of the thesaurus from a lexico-semantic resource to an encyclopaedia. It is an editorial decision whether to include proper names, but for the purposes of my proposed compilation, it has been found theoretically and practically unsound to include them.

Inclusion of obsolete usages

Ideally, one would wish to omit obsolete lexical items from a thesaurus purportedly of modern Tamil. Decisions in this area are, however, governed by the sociolinguistic context of the language. There is an influential movement in Tamil advocating its purification by stripping away borrowings, especially from Sanskrit and English, and replacing them with native creations as well as resurrected obsolete words. See Annamalai (1979) and Britto (1986) for extensive detailed discussions of the purist movement in Tamil. Promoters of purism argue that the intake of a large amount of lexis from foreign sources detract from the beauty and independence of native Tamil words and will lead to a degeneration of the language. Their opponents' arguments centre on the fact that the native stock is insufficient for the creation of adequate lexis to express *inter alia* concepts in contemporary scientific and technical fields. Since 1967, the purist movement has had the active support of the Tamil Nadu State Government, and the governmentally-controlled language planning process has insisted upon the use of pure Tamil vocabulary in e.g. school textbooks, administrative writing and the coining of scientific and technical terminology. Thus even in primary-level textbooks, one finds such previously obsolete items as அகவை 'age' and ஞமலி 'dog'. There are some linguists who are worried that purism will widen the gap between the diglossic varieties of Tamil to the extent that H will have to be taught as a separate language; but Annamalai's (1979: 51) words offer some encouragement on this score:

> There [are] two trends, viz. purification and simplification, in the efforts of modernization of Tamil in [India]. Present day written Tamil is coming closer to the colloquial in sandhi and syntax and to some extent in morphology – a move in the opposite direction of archaism ... It may be said that purification as opening to archaic sources (i.e. classicalization) has succeeded to a large extent in spelling and vocabulary, but simplification has succeeded in sandhi and syntax. In morphology, the morphological structures are closer to the colloquial with new morphemes and their combinations, but the phonological shapes of the morphemes, which follow archaic spelling, are removed from colloquial pronunciation.

In the context of modernisation of Tamil, a thesaurus will have to include such resurrected words from classical sources as have regained currency in the modern language.

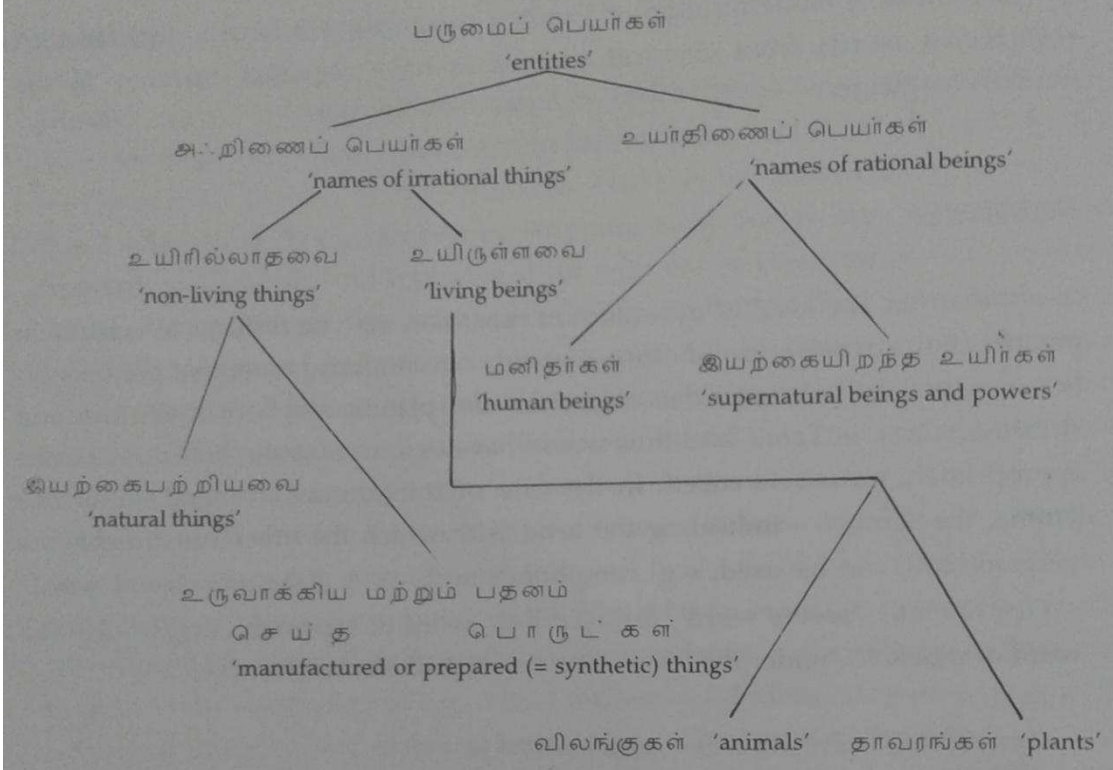
Repetition

In a thesaurus, the ideal of avoidance of repetition may be difficult to achieve in practice, but extensive use of cross-reference can minimise repetitive occurrences For example, lexical items denoting stars and planets are both NATURAL and SUPERNATURAL in Tamil, and thus would have to appear under both heads or be appropriately cross-referenced. In the case of compounds arranged under one lemma, the symbol ~ indicating the head with which the other constituents are compounded, can be used, e.g. compounds such as நற்சொல் 'good word', கடுஞ்சொல் 'severe word', சுடுசொல் 'word of reproach', பழிச்சொல் 'word of rebuke'... under the lemma சொல் 'word' can be given as:

சொல் 'word' (நற்--, கடுஞ்--, சுடு--, பழிச்-- ...

Presentation

The presentation in the thesaurus under consideration will include superordinate terms (given a serial number order according to the hierarchy established); headwords of a group of co-lexical items are underlined. Synonyms are separated by commas, and non-synonyms by semi-colons. A superordinate term is separated from its co-hyponyms by a colon. The modified hierarchical scheme of classification of entities is given below.



Sample (Rajendran 1991)

அ	பருமைப் பெயர்கள் 'names of concrete things, i.e. entities'
1	அ·றிணைப் பெயர்கள் 'names of irrational things'
1.1.	உயிரில்லாதவை 'non-living things'
1.1.1.	இயற்கை பற்றியவை 'natural things'
1.1.1.1	பூகோளம் பற்றியவை 'geographical things'
1.1.1.1.1	விண்வெளி பற்றியவை 'celestial-atmospheric things'

க, வானம் 'sky', ஆகாயம் 'sky', வான் 'sky', ஆகாசம் 'sky', விண் 'sky', விசும்பு 'sky', அந்தரிட்சம் 'sky', அந்தரம் 'sky'; கீழ்வானம் 'lower sky'; மேல்வானம் 'upper sky'

ங ...

1.1.1.1.2 இயற்கைக்கு அப்பாற்பட்டவை 'supernatural things'
1.1.1.2 இயற்கைப் பொருட்கள் 'natural substances'
1.1.2 உருவாக்கிய மற்றும் பதனம் செய்த பொருட்கள் 'manufactured or prepared things'
1.2 உயிருள்ளவை 'living beings'
1.2.1 தாவரங்கள் 'plants'
1.2.2 விலங்குகள் 'animals'
2 உயர்திணைப் பெயர்கள் 'names of rational beings'

ஆ நிகழ்வுகள் 'events'
1 பௌதிக நிகழ்வுகள் 'physical events'
1.1 பொது 'general'
நிகழ்தல் 'happening', சம்பவித்தல் 'happening', விளைதல் 'happening', நேர்தல் 'happening', பலித்தல் 'happening'; நிறைவேறுதல் 'accomplishing';
நிகழ்ச்சி 'incident', சம்பவம் 'incident'; விளைவு 'result'
1.2 சூரியன் சந்திரன் சம்பந்தப்பட்ட நிகழ்வுகள் 'events relating to the sun and moon'
1.2.1 விடிதல் 'dawning', புலர்தல் 'dawning'; இருட்டுதல் 'darkening';
1.2.2 ...
2 உடற்கூறு சம்பந்தப்பட்ட நிகழ்வுகள் 'physiological events';
3 புலனுணர்வுகள் 'sensory';
4 உணர்ச்சிகள் 'emotions';
5 அறிவு சம்பந்தப்பட்ட நிகழ்வுகள் 'intellectual events'
6 கருத்துப் பரிமாற்றம் 'communication'
7 சேர்க்கை 'association'
8 கட்டுப் படுத்துகை 'control';
9 சலனம் 'movement'
10 தாக்கம் 'impact'
11 கைமாற்றம் 'transfer'
12 பல்கூட்டான செயல்கள் 'complex activities'

இ அருவப் பெயர்கள் 'abstracts'
1 நேரம் 'time';
2 தூரம் 'distance';

3 கொள்ளளவு 'volume';
4 வேகம் 'velocity';
5 சூடு 'temperature';
6 நிறம் 'colour';
7 எண் 'number'
8 தகுநிலை 'status';
9 பரிசுத்தம் 'purity';
10 கவர்ச்சி 'attractiveness';
11 வயது 'age';
12 உண்மை பொய் 'truth–falsehood';
13 நல்லது கெட்டது 'good–bad';
14 ஆற்றல் 'capacity';
15 உடல் நலை 'state of health'
...

ஈ தொடர்பன்கள் 'relationals'
1 இடம் சம்பந்தப்பட்டவை 'spatial';
2 காலம் சம்பந்தப்பட்டவை 'temporal';
3 சுட்டிக்காட்டுபவை 'deictic';
4 வருமுறையானவை 'logical'; ...

Conclusion

A thesaurus aims to classify words by concepts, topics or subjects. Nida's classification of words by concepts offers an insight into the semantic structure of a language. The *nikaṇṭus* endeavoured to classify the vocabulary of Tamil, but evidently without the foundation of any strong structural theory.

A word acquires its referential meaning in being a member of a semantic domain by the common features it shares with other members in that domain, and having contrasting features which separate it from other members of the domain. Thus it is the semantic relations among words, such as synonymy, hyponymy, compatibility and incompatibility, which help one to organise words in semantic domains in a hierarchical fashion. This structuring of the vocabulary of a language into concepts and domains leads to a thesauric classification, that is

one which can largely be defined as *entities, events, abstracts* and *relationals*, underlying the syntactic categories of *nouns, verbs, adjectives/adverbs* and *prepositions/postpositions* respectively.

Though a universal semantic classification of words on a conceptual basis is probably possible, there need to be changes at the lower levels of classification to accommodate those features peculiar to any particular language. The thesauric classification of Tamil, for example, will not only show how the vocabulary of Tamil can be structured, but will also focus our attention on the gaps in the vocabulary system. The structuring of the vocabulary of Tamil into semantic domains is one of the principal problems here, but there are others such as finding superordinate terms, adopting scientific taxonomy to folk taxonomy, bringing in the part-whole relations among items, the selection of stylistic alternants and inflectional formations, the inclusion or exclusion of loans from English and elsewhere, phrases and compounds, proper names, obsolete usages, and the avoidance of repetitions, which the thesaurus compiler has to address in order to make judgements on defined theory. The presentation of Tamil lexis in a structural framework has to fulfil the expectations of a thesaurus; it has also to be undertaken in such a way that readers derive the maximum benefit from it.

References

Lexicographical

Rajendran, S. (1991) தற்காலத் தமிழ்ச் சொற்களஞ்சியம். Unpublished MS, Tamil University, Thanjavur.

Sadasivam, M. (n.d.) A DICTIONARY OF EPITHETS AND NAMES OF GOD. Unpublished MS.

Sadasivam, M. (n.d.) A DICTIONARY OF EPITHETS AND NAMES OF GODDESS UMA. Unpublished MS.

Sadasivam, M. (n.d.) A DICTIONARY OF EPITHETS AND NAMES OF LORD MURUGA. Unpublished MS.

Sadasivam, M. (n.d.) A DICTIONARY OF EPITHETS AND NAMES OF LORD SIVA. Unpublished MS.

Sadasivam, M. (n.d.) A DICTIONARY OF EPITHETS AND NAMES OF LORD VISHNU. Unpublished MS.

Sivasubrahmanya Kavirayar (1810) NAMATIPAM.

Tucker, C. (1986) A DICTIONARY OF ENGLISH LOAN WORDS IN MODERN TAMIL. New Delhi: Biblia Impex.

Vaiyapuri Pillai, S. (ed.) (1930) NAMATIPA NIKANTU. Madras: B.N. Press.

Other

Annamalai, E. (1979) 'Movement for linguistic purism: the case of Tamil.' In E. Annamalai (ed.), *Language Movements in India*, pp. 35–59. Manasagangotri: Central Institute of Indian Languages.

Asher, R.E. (1985) *Tamil*. London: Croom Helm.

Britto, F. (1986) *Diglossia: a study of the theory with application to Tamil*. Washington, DC: Georgetown University Press.

Cruse, D.A. (1986) *Lexical Semantics*. Cambridge: Cambridge University Press.

Deiva Sundaram, N. (1981) *Tamil Diglossia*. Tirunelveli: Nainar Pathippagam.

James, G. (1991) *Tamil Lexicography*. Tübingen: Niemeyer (Lexicographica Series Maior 40).

Jones, K.S. (1986) *Synonymy and Semantic Classification*. Edinburgh: Edinburgh University Press.

Lyons, J. (1977) *Semantics*. Vol. 1. Cambridge: Cambridge University Press.

Nida, E.A. (1975a) *Componential Analysis of Meaning: an introduction to semantic structure*. The Hague: Mouton.

Nida, E.A. (1975b) *Exploring Semantic Structures*. Munich: Fink.

Rajendran, S. (1978) 'Syntax and semantics of Tamil verbs.' PhD thesis, University of Poona.

Rajendran, S. (1981) 'Semantic structure of Tamil verbs.' ஆய்வுக்கோவை (13th All India Tamil Teachers' Association Conference, Vol. 2), pp. 305–310.

de Saussure, F. (1916) *Cours de linguistique générale*. Paris: Payot (English translation, *A Course in General Linguistics*. New York: Philosophical Library).

Schiffman, H. (1978) 'Diglossia and purity/pollution in Tamil.' In C. Maloney (ed.), *Language and Civilization Change in South Asia*, pp. 98–110. Leiden: Brill (Contributions to Asian Studies 11).

Sethu Pillai, R.P. (1974) *Tamil – Literary and Colloquial*. Madras: University of Madras (Tamil Department Publication 35).

Shanmugam Pillai, M. (1972a) 'Tamil – literary and colloquial.' *Linguistics*, 81: 83–91.

Shanmugam Pillai, M. (1972b) 'Tamil today.' *Indian Linguistics*, 33: 67–71.

de Silva, M.W.S. (1976) *Diglossia and Literacy*. Manasagangotri: Central Institute of Indian Languages (Adult Literacy Series 2).

Ullmann, S. (^{2}1959) *The Principles of Semantics*. Oxford: Blackwell.

Vadivelan, R. (1980) மு.ச. சிவம் வாழ்க்கைக் குறிப்புக்களும் தமிழுக்குச் செய்த ஆக்கப்பணிகளும். Erode: Mira Nilaiyam.